

MSX2 mediates entry of human pluripotent stem cells into mesendoderm by simultaneously suppressing SOX2 and activating NODAL signaling

Qingqing Wu^{1,2}, Leisheng Zhang^{1,2}, Pei Su^{1,2}, Xiaohua Lei³, Xin Liu^{1,2}, Hongtao Wang^{1,2}, Lisha Lu⁴, Yang Bai^{1,2}, Tao Xiong⁴, Dong Li⁵, Zhengmao Zhu⁶, Enkui Duan³, Erlic Jiang^{1,2}, Sizhou Feng^{1,2}, Mingzhe Han^{1,2}, Yuanfu Xu^{1,2}, Fei Wang⁷, Jiayi Zhou^{1,2}

¹State Key Laboratory of Experimental Hematology, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 288 Nanjing Road, Tianjin 300020, China; ²Center for Stem Cell Medicine, Chinese Academy of Medical Sciences & Department of Stem Cells and Regenerative Medicine, Peking Union Medical College, Beijing, China; ³State Key Laboratory of Reproductive Biology, Institute of Zoology, CAS, Beijing 100101, China; ⁴College of Life Sciences at Yangtze University, Jingzhou, Hubei 434025, China; ⁵Department of Oncology, Shanghai Third People's Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai 201900, China; ⁶College of Life Sciences, Nankai University, Tianjin 300071, China; ⁷Department of Cell and Developmental Biology and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

How BMP signaling integrates into and destabilizes the pluripotency circuitry of human pluripotent stem cells (hPSCs) to initiate differentiation into individual germ layers is a long-standing puzzle. Here we report muscle segment homeobox 2 (MSX2), a homeobox transcription factor of msh family, as a direct target gene of BMP signaling and a master mediator of hPSCs' differentiation to mesendoderm. Enforced expression of MSX2 suffices to abolish pluripotency and induce directed mesendoderm differentiation of hPSCs, while MSX2 depletion impairs mesendoderm induction. MSX2 is a direct target gene of the BMP pathway in hPSCs, and can be synergistically activated by Wnt signals via LEF1 during mesendoderm induction. Furthermore, MSX2 destabilizes the pluripotency circuitry through direct binding to the SOX2 promoter and repression of SOX2 transcription, while MSX2 controls mesendoderm lineage commitment by simultaneous suppression of SOX2 and induction of NODAL expression through direct binding and activation of the *Nodal* promoter. Interestingly, SOX2 can promote the degradation of MSX2 protein, suggesting a mutual antagonism between the two lineage-specifying factors in the control of stem cell fate. Together, our findings reveal crucial new mechanisms of destabilizing pluripotency and directing lineage commitment in hPSCs.

Keywords: MSX2; hPSC; SOX2; NODAL; mesendoderm

Cell Research (2015) 25:1314–1332. doi:10.1038/cr.2015.118; published online 2 October 2015

Introduction

Human pluripotent stem cells (hPSCs), including human embryonic stem cells (hESCs) derived from the inner cell mass (ICM) or human induced pluripotent stem

cells reprogrammed from somatic cells, can self-renew almost indefinitely in culture and have the remarkable potential to differentiate into nearly all cell types in the human body [1, 2]. They therefore hold great promise for developmental studies, drug screening, cell-based therapy and disease modeling. Harnessing the full potential of hPSCs requires a deeper understanding of the signaling mechanisms governing pluripotency and directed differentiation into the early embryonic lineages [3, 4].

Significant advances have been made towards the understanding of the extrinsic growth factors, intracellular

Correspondence: Jiayi Zhou

Tel: +86-22-23909412; Fax: +86-22-23909412.

E-mail: zhoujx@ihcams.ac.cn

Received 13 November 2014; revised 13 July 2015; accepted 10 August 2015; published online 2 October 2015

pathways and nuclear factors united to control pluripotency and self-renewal of hPSCs. Basic FGF (bFGF), IGF and TGF- β /Activin/Nodal cooperate to support pluripotency by stabilizing the core transcriptional circuitry consisting of OCT4 (also known as POU5F1), SOX2 and NANOG, which function in concert to positively regulate target genes necessary for pluripotency and to repress a variety of lineage specification factors [5-8]. Signals from the extrinsic factors are integrated and transmitted by cytoplasmic pathways including PI3K, MAPK/ERK, Smads and mammalian target of rapamycin (mTOR), which support the undifferentiated state of hPSCs [6, 9-11].

In contrast to our understanding of the factors that govern pluripotency, molecular program that controls hPSC exit from pluripotency and entry into the embryonic germ layers is still poorly defined. The Wnt/ β -catenin and BMP signaling pathways are critical for hPSC differentiation into the early embryonic lineages. Activation of Wnt signaling results in a loss of pluripotency and drives hPSC differentiation towards endoderm and mesoderm [12-14], and activation of BMP signaling leads to mesendoderm or trophoectoderm differentiation, depending on the dose and duration of stimulation [15, 16]. Interestingly, Nodal/Activin signaling cooperates with bFGF to maintain pluripotency [17, 18], but promotes mesendoderm differentiation of hESCs in the absence of bFGF [19-21]. Despite these studies, there are still significant gaps in the knowledge of the signaling mechanisms that regulate cell fate transitions from pluripotency to the embryonic germ layers. Especially, the intracellular and nuclear factors that mediate hPSC mesendoderm differentiation, the direct target gene(s) of BMP signaling that mediate BMP's function in mesendoderm induction, and how they interact with and destabilize the core pluripotency circuitry and induce mesendoderm lineage commitment remain to be defined.

By taking advantage of a high-efficiency neural induction model and large-scale gene profiling analysis, we previously identified muscle segment homeobox (MSX2) as a responsive gene to BMP stimulation in hPSCs [22]. This finding was subsequently confirmed in a separate study [23]. MSX2, a homeobox-containing transcription factor, belongs to the highly conserved and widely expressed msh family [24-26]. MSX2 has been described as a transcription repressor, but emerging evidence suggests that it can also activate downstream target genes [27, 28]. Experiments in mouse model have revealed an essential function for *Msx2* in craniofacial, limb and ectodermal organogenesis — *Msx2* deletion mutations result in profound defects in the development of skull vault, tooth, hair follicle and mammary gland [26, 29].

In keeping with the defects in mice, mutations of MSX2 are associated with Boston-type craniosynostosis and parietal foramina [30-33]. The function of MSX2 in craniofacial, limb and mammary gland development is linked to its ability to regulate epithelial to mesenchymal transition [26, 34-36]. Despite the knowledge of MSX2 involvement in BMP signaling and organogenesis, the role of MSX2 in early embryonic development, especially in human, remains to be elucidated.

In this study, we explored the function of MSX2 in hPSC fate determination, revealing an essential role of MSX2 in hPSCs' exit from pluripotency and entry to mesendoderm. MSX2 is both necessary and sufficient for mesendoderm differentiation of hPSCs. MSX2 acts as a direct target gene of the BMP pathway in hPSCs, and it can be synergistically activated by Wnt signals via LEF1 during mesendoderm differentiation. Furthermore, MSX2 destabilizes the pluripotency circuitry through direct binding to the *SOX2* promoter and repression of *SOX2* transcription, while MSX2 induction of mesendoderm differentiation requires simultaneous suppression of *SOX2* and activation of Nodal signaling. Interestingly, *SOX2* does not merely lie downstream of MSX2 but can promote the MSX2 protein degradation, suggesting a mutual antagonism between these two factors in the control of stem cell fate.

Results

Enforced MSX2 expression induces directed hESC mesendoderm differentiation

To explore the function of MSX2 in fate determination of hPSCs, we overexpressed MSX2 in hESCs using a previously described doxycycline (DOX) inducible lentiviral expression system and assessed its effect [37]. We used a GFP-MSX2 fusion gene which allowed us to monitor its expression in hESCs in real time (Supplementary information, Figure S1A). As expected, GFP expression was largely undetectable in the absence of DOX but could be readily seen 24 h after DOX was added (Supplementary information, Figure S1B). A high percentage of GFP-MSX2-positive cells were detected after colony isolation and drug selection ($90.8\% \pm 5.1\%$; Supplementary information, Figure S1B).

MSX2 overexpression induced profound morphological changes in hESCs. 72 h after DOX was added, hESCs began to flatten and spread out. After 120 h, the colony integrity of hESCs was completely abolished; instead, large flat cells formed a uniform layer (Figure 1A). The alterations in hESC morphology suggested an induction of differentiation. Indeed, real-time PCR analysis revealed a rapid downregulation of pluripotency

marker SOX2, while expression of POU5F1/OCT4 and NANOG, which was unaltered or moderately elevated at 24 h, decreased gradually (Figure 1B). Concomitant with the downregulation of pluripotency markers, expression of mesendoderm markers T (also known as BRACHYURY) and MIXL1 increased dramatically, peaking at 72 h after DOX addition (Figure 1B). In contrast, neuroectoderm markers PAX6 and SOX1 were substantially downregulated (Figure 1B). The effect of MSX2 overexpression on pluripotency and differentiation marker expression was confirmed at the protein level by western blotting and immunofluorescence analysis (Figure 1C; Supplementary information, Figure S1C).

Strikingly, T was found in nearly all GFP-MSX2-overexpressing cells, while no PAX6 and SOX1 expression was detected (Figure 1C). Furthermore, GFP-MSX2-overexpressing hESCs could no longer form teratomas *in vivo*, in sharp contrast to cells that overexpressed GFP only (Figure 1D), indicating that hESCs with MSX2 overexpression lost the potential to differentiate into three germ layers, a hallmark of pluripotency. The effects of MSX2 overexpression were observed in H1 hESCs cultured in mTeSR1 medium, E8 medium, and mouse embryonic fibroblast (MEF)-conditioned medium (CM), suggesting conserved MSX2 effect under various culture conditions (Supplementary information, Figure S1C). Furthermore,

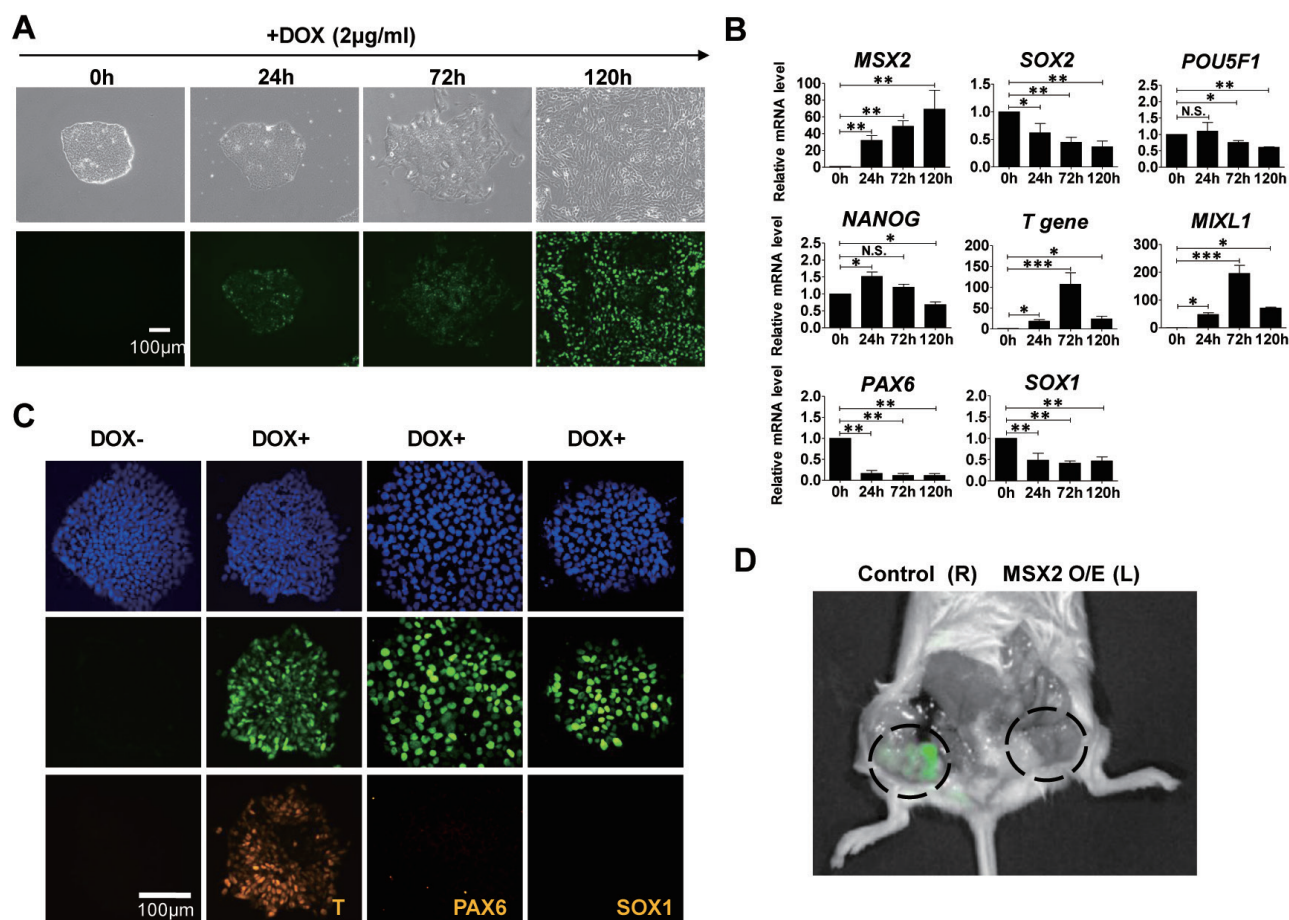


Figure 1 MSX2 suffices to induce hESC mesendoderm differentiation. **(A)** Phase contrast (top) and fluorescence (bottom) images of GFP-MSX2 H1 hESCs after addition of DOX (2 μ g/ml). A time-course analysis of the same colony is shown (from 0 h to 120 h). Scale bar, 100 μ m. **(B)** mRNA levels of MSX2, pluripotency and lineage-specific genes assessed by real-time PCR in H1 hESCs with MSX2 overexpression induced by DOX addition at different time points. All values are normalized to the level (= 1) of mRNA in the cells before adding DOX (0 h). Results are shown as means \pm SEM ($n = 5$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS, not significant. **(C)** Immunofluorescence of T, PAX6 and SOX1 proteins (orange) at 72 h in H1 hESCs cultured as monolayer with or without GFP-MSX2 overexpression. Scale bar, 100 μ m. **(D)** Teratoma formation of hESCs in SCID mice. GFP H1 hESCs (control) and GFP-MSX2 H1 hESCs were injected to the right and left hind legs, respectively. Teratomas and GFP expression were only detected in the right hind legs (See also Supplementary information, Figure S1).

MSX2 overexpression also induced mesendoderm differentiation in H9 hESCs, suggesting conserved responses in hPSCs (Supplementary information, Figure S1D). Thus, enforced MSX2 expression suffices to abolish pluripotency and induce directed mesendoderm differentiation in hPSCs.

Previous report that MSX1 can respond to BMP signaling activation [38, 39], led us to also assess the effect of MSX1 overexpression in hESCs (Supplementary information, Figure S1E). While profound morphological changes could also be observed 120 h after DOX addition, the upregulation of *T* and *MIXL1* mRNA levels was much lower than that caused by MSX2 overexpression (Supplementary information, Figure S1F). Moreover, enforced expression of MSX1 did not repress expression of neuroectoderm markers such as PAX6 and SOX1 (Supplementary information, Figure S1F). These results indicate that MSX2 is much more potent than MSX1 in inducing directed mesendoderm differentiation of hESCs.

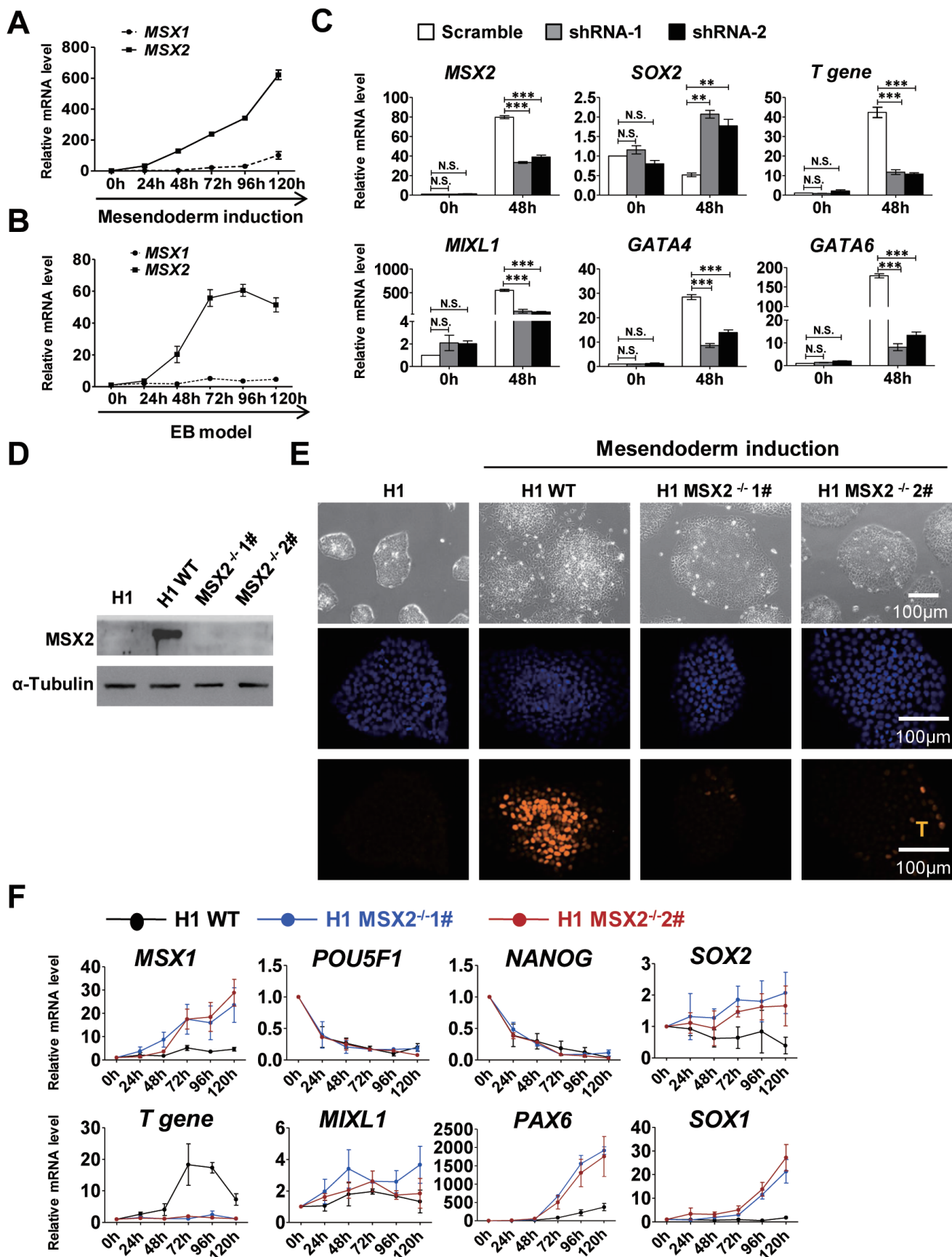
MSX2 is required for hPSCs' exit from pluripotency and entry to mesendoderm lineage

We next asked whether MSX2 was required for mesendoderm differentiation of hPSCs. We induced directed mesendoderm differentiation using a previously described protocol with some modifications [40] (Supplementary information, Figure S2A). The presence of Activin A, BMP4, Wnt3a and bFGF induced H1 hESCs to adopt a differentiation morphology (Supplementary information, Figure S2B) and increased the expression of mesendoderm markers including *T*, *MIXL1* and others (Supplementary information, Figure S2C). Time-course analysis revealed a rapid, time-dependent upregulation of both *MSX2* mRNA and protein upon induction of differentiation (Figure 2A, Supplementary information, Figure

S2D). *MSX1* was also upregulated, but the increase was much less and slower than *MSX2*. During spontaneous differentiation of hPSCs induced via embryoid body (EB) formation, *MSX2* expression was also upregulated within 24–48 h and peaked at 72–96 h. In contrast, *MSX1* expression remained at a very low level up to 120 h (Figure 2B). Thus, we mainly focused on *MSX2* in the rest of the study.

We first used small hairpin RNAs (shRNAs) to deplete *MSX2* in hESCs and determined the impact on mesendoderm differentiation. *MSX2* depletion in H1 hESCs under self-renewal condition (i.e., in mTeSR1 medium) had minimal effect, presumably due to the low expression level of *MSX2* (Figure 2C-0 h, Supplementary information, Figure S3A-0 h). Strikingly, upon induction of mesendoderm differentiation, cells depleted of *MSX2* exhibited much lower levels of *T* mRNA and protein than control cells (Figure 2C, Supplementary information, Figure S3B). In addition, expression of other mesendoderm markers such as *MIXL1*, *GATA4* and *GATA6* was also significantly reduced (Figure 2C). In contrast, levels of pluripotency marker *POU5F1/OCT4*, neuroectoderm markers *PAX6* and *SOX1*, and neuroectoderm/pluripotency marker *SOX2* were substantially elevated (Supplementary information, Figure S3A and S3B). Impairment of mesendoderm differentiation upon *MSX2* depletion was also observed in H9 hESCs (Supplementary information, Figure S3C and S3D). Unexpectedly, the pluripotency marker *NANOG* was repressed upon *MSX2* depletion in both H1 and H9 hESCs (Supplementary information, Figure S3A-S3D). We speculated that this might be related to *NANOG* function in suppressing neural differentiation and in patterning different subtypes of mesoderm cells after exit of hESCs from pluripotency, as described previously [41, 42]. To further confirm the

Figure 2 *MSX2* is essential for hESC mesendoderm specification. **(A, B)** Time-course analysis of *MSX1* and *MSX2* expression during mesendoderm differentiation of H1 hESCs cultured as monolayer **(A)** and during spontaneous differentiation in EB model **(B)** assessed by real-time PCR. All values are normalized to the level (= 1) of mRNA in the cells cultured in mTeSR1 medium before differentiation was induced (0 h). Results are shown as means \pm SEM ($n = 3$). **(C)** Real-time PCR analysis of H1 hESCs depleted of *MSX2* by shRNA-1 or shRNA-2 or expressing a scramble shRNA (Scramble) before (0 h) and 48 h after mesendoderm induction. All values are normalized to the level (= 1) of mRNA in the cells infected with scramble shRNA lentivirus and cultured in mTeSR1 before mesendoderm differentiation was induced. Results are shown as means \pm SEM ($n = 3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS, not significant. **(D)** Western blotting analysis confirms the deletion of *MSX2* in two knockout H1 hESC lines (*MSX2*^{-/-} 1# and *MSX2*^{-/-} 2#). Wild-type (WT) and mutant H1 hESC cells were cultured in mesendoderm induction condition for 48 h; H1 hESCs cultured in mTeSR1 served as negative control (H1). α -Tubulin was used as a loading control. A typical experiment from three separate experiments is shown. **(E)** Phase contrast and fluorescence images of *T* expression in H1 cultured in mTeSR1 and in WT and *MSX2*-deleted H1 hESC cells cultured in mesendoderm differentiation induction condition for 48 h. Scale bar, 100 μ m. **(F)** Time-course analysis of gene expression in WT H1 and in *MSX2*-deleted H1 cells (*MSX2*^{-/-} 1#; *MSX2*^{-/-} 2#) during spontaneous EB differentiation by real-time PCR. All values are normalized to the level (=1) of mRNA in the cells cultured in mTeSR1 before differentiation (0 h). Results are shown as means \pm SEM ($n = 3$; see also Supplementary information, Figures S2-S4).



role of *MSX2* in hESC early differentiation and exclude the potential ambiguity brought by shRNAs, we deleted *MSX2* gene in hESCs using CRISPR-CAS9 technology. We designed two sgRNAs each targeting a separate exon in human *MSX2* gene, using a previously described method [43] (Supplementary information, Figure S4A). Gene sequencing and western blotting analysis confirmed the establishment of two homozygous H1 hESC lines with *MSX2* deletion (*MSX2*^{-/-} 1# and *MSX2*^{-/-} 2#; Figure 2D, Supplementary information, Figure S4B). We induced both directed and spontaneous mesendoderm differentiation of these two knock-out hESC lines. Consistent with the results by using shRNAs, *MSX2* deletion significantly reversed the differentiation morphology of hESCs and inhibited mesendoderm marker expression (Figure 2E). Furthermore, compared with the EBs derived from wild-type hESCs, *MSX2*-deleted EBs had a much low level of *T* expression, while neuroectoderm markers *PAX6* and *SOX1* and neuroectoderm/pluripotency marker *SOX2* were substantially upregulated (Figure 2F). These results suggest that EBs with *MSX2* deletion has a strong bias toward neuroectoderm differentiation at the expense of mesendoderm differentiation. Interestingly, in *MSX2*-deleted EBs, *MSX1* level was enhanced, presumably due to compensatory mechanism [25, 44, 45] (Figure 2F). Together, these findings confirm the essential role of *MSX2* in hPSC mesendoderm differentiation.

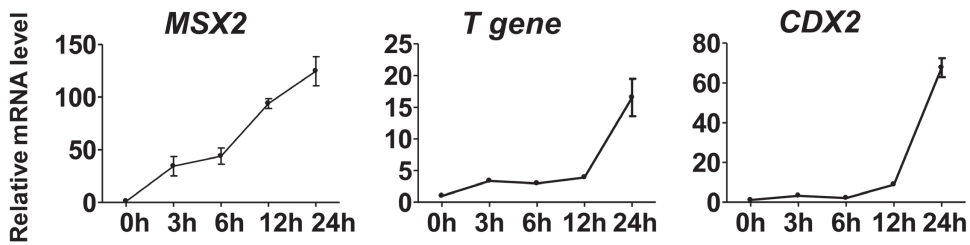
MSX2 is a direct target of BMP signaling in hESCs

MSX2 is a component of multiple developmental pathways including BMP, Wnt and FGF [23, 27, 46, 47]. During mesendoderm differentiation of hESCs, *MSX2* is rapidly upregulated in the presence of Activin A, BMP4, Wnt3a and bFGF, leading us to ask which of the extrinsic factors could induce *MSX2* expression. As expected, stimulation of hESCs with BMP4 led to a rapid and dramatic upregulation of *MSX2* mRNA within 3 h of BMP4 stimulation (Figure 3A and Supplementary information, Figure S4C). Furthermore, induction of *MSX2* expression consistently preceded *T* and *CDX2* — two previously reported BMP4 targets in hESCs [48] (Figure 3A and Supplementary information, Figure S4C). In contrast, treatment of cells with Activin A, Wnt3a, bFGF and other factors caused little effect on *MSX2* and *MSX1* expression (Figure 3B). Interestingly, combination BMP4 and Wnt3a led to a synergistic stimulation of *MSX2* expression, while the combination of BMP4 with other factors failed to cause additive effect (Figure 3C).

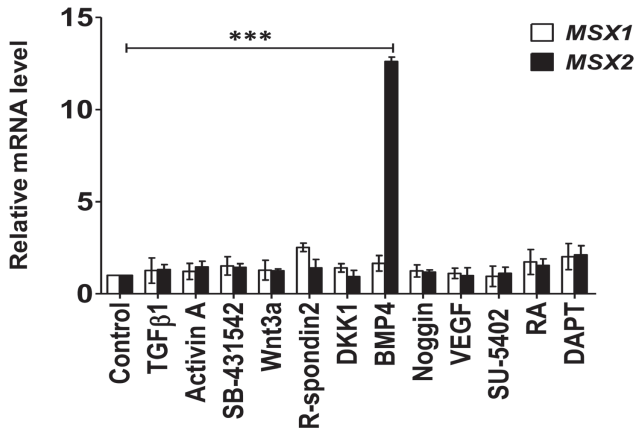
It was previously reported that *MSX2* promoter contains a 52 bp phylogenetically conserved BMP-restricted Smad-binding element termed BMP-responsive element (BMPRe; Supplementary information, Figure S4D) [49, 50]. Binding of Smads such as Smad1 to this site has been observed in mouse cells [49]. We conducted chromatin immunoprecipitation coupled to detection by quan-

Figure 3 *MSX2* is a direct downstream target of the BMP pathway. **(A)** Time-course analysis of *MSX2*, *T* and *CDX2* mRNA levels in H1 hESCs with BMP4 stimulation (20 ng/ml; from 0–24 h) by real-time PCR. All values are normalized to the level (= 1) of mRNA in the cells cultured in mTeSR1 before stimulation (0 h). Results are shown as means ± SEM (*n* = 3). **(B)** Real-time PCR analysis of *MSX1* and *MSX2* mRNA levels in H1 hESCs treated with various soluble factors and inhibitors for 12 h. Concentration of each factors are: TGFβ1 20 ng/ml; Activin A 20 ng/ml; SB-431542 20 μM; Wnt3a 20 ng/ml; R-spondin2 50 ng/ml; DKK1 250 ng/ml; BMP4 20 ng/ml; Noggin 300 ng/ml; VEGF 20 ng/ml; SU-5402 20 μM; RA 20 μM; DAPT 20 μM. All values are normalized to the level (= 1) of mRNA in the cells cultured in mTeSR1 without treatment. Results are shown as means ± SEM (*n* = 3). ****P* < 0.001. **(C)** Real-time PCR analysis of *MSX2* mRNA level in H1 hESCs treated with BMP4 alone or BMP4 combined with other factors for 12 h. Concentration of each factors are: BMP4 20 ng/ml; Wnt3a 20 ng/ml; Activin A 20 ng/ml; VEGF 20 ng/ml; TGFβ1 20 ng/ml; bFGF 10 ng/ml; Nodal 200 ng/ml; RA 20 μM. All values are normalized to the level (= 1) of mRNA in the cells cultured in mTeSR1 without treatment. Results are shown as means ± SEM (*n* = 3). **P* < 0.05. **(D)** ChIP-qPCR analysis of BMP-responsive element (BMPRe) on *MSX2* promoter in H1 hESCs treated with BMP4 (20 ng/ml) for 6 h. Non-specific IgG was used as isotype control. All values are normalized to that of their corresponding input samples. Results are shown as means ± SEM (*n* = 3). ****P* < 0.001; NS, not significant. **(E)** Relative luciferase activity in H1 cells transfected with PGL4.2 basic luciferase construct or the pGL4.2 construct containing BMPRe from *MSX2* promoter (pBMPRe-LUC) 6 h after BMP4 treatment with a dose gradient. All values are normalized to the level (= 1) of luciferase activity in cells transfected with the pGL4.2 basic vector. Results are shown as means ± SEM (*n* = 3). **(F)** Real-time PCR analysis of *LEF1* and *MSX2* mRNA level in *LEF1*-overexpressing H1 hESCs with or without BMP4 (20 ng/ml) treatment. All values are normalized to the level (= 1) of mRNA in the cells cultured in mTeSR1 without *LEF1* overexpression and BMP4 treatment. Results are shown as means ± SEM (*n* = 3). **P* < 0.05; ***P* < 0.01; ****P* < 0.001; NS, not significant. **(G)** Relative luciferase activity in 293T cells transfected with PGL4.2 basic luciferase construct or pGL4.2 construct containing BMPRe of *MSX2* promoter (pBMPRe-LUC). Expression vectors for human SMAD1, SMAD4 and *LEF1* were co-transfected with pBMPRe-LUC as described in the graph. All values are normalized to the level (= 1) of the luciferase activity in cells transfected with pGL4.2 basic vector. Results are shown as means ± SEM (*n* = 3). **P* < 0.05; ***P* < 0.01; ****P* < 0.001 (see also Supplementary information, Figure S4).

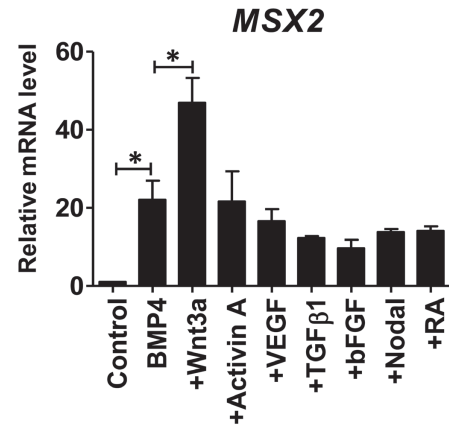
A



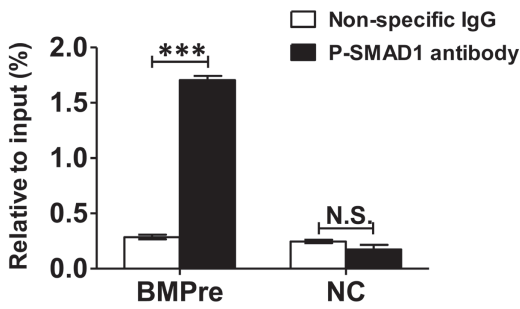
B



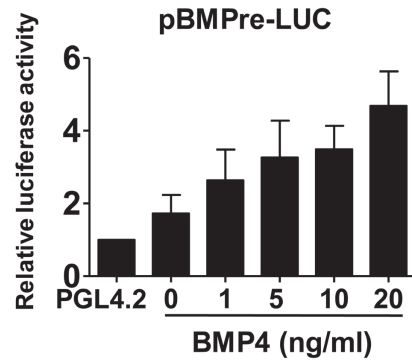
C



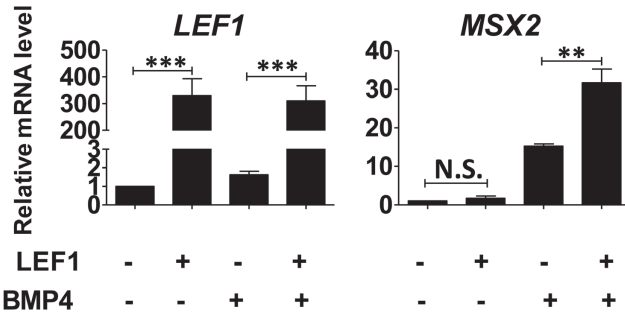
D



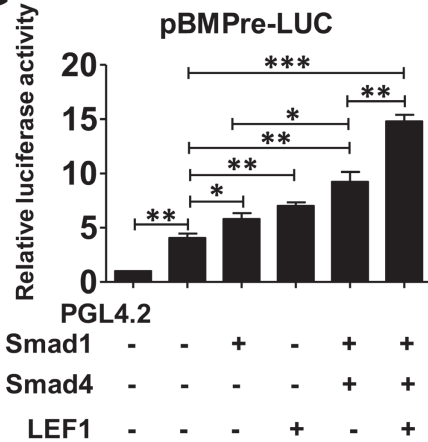
E



F



G



titative real-time PCR (ChIP-qPCR) to examine whether BMP-restricted SMADs were capable of binding to BMPre in hESCs. Indeed, our results indicated there was strong binding of phosphorylated (activated) Smad1 to this site in hESCs (Figure 3D). To further explore whether BMP activation of MSX2 requires a binding to the Smad1 binding sites of BMPre within *MSX2* promoter, we isolated the 2 kb fragment containing BMPre within the *MSX2* promoter and tested its response to BMP stimulation using a luciferase-based reporter assay. We found that BMP stimulation enhanced the activity of BMPre in hESCs in a dose-dependent manner (Figure 3E). Thus, BMP signaling directly activates MSX2.

We have shown that BMP and Wnt activation can synergistically activate MSX2. It has been previously reported that LEF1, a transcriptional mediator of Wnt signaling, interacts with Smad1 and Smad4 to regulate graded expression of *Msx2* in mouse ESCs [51]. We thus tested whether LEF1 was also involved in induction of MSX2 expression by Wnt signaling in hESCs. We found enforced expression of LEF1 enhanced MSX2 expression synergistically with BMP treatment (Figure 3F). Moreover, the luciferase activity of BMPre within the *MSX2* promoter could be further enhanced by co-transfection of Smad1, Smad4 and LEF1 (Figure 3G). Thus, LEF1 acts as a key factor that mediates the function of Wnt signaling in MSX2 activation.

Taken together, our data suggest that MSX2 serves as a master mediator of hPSCs exit from pluripotency and entry to the mesendoderm fate.

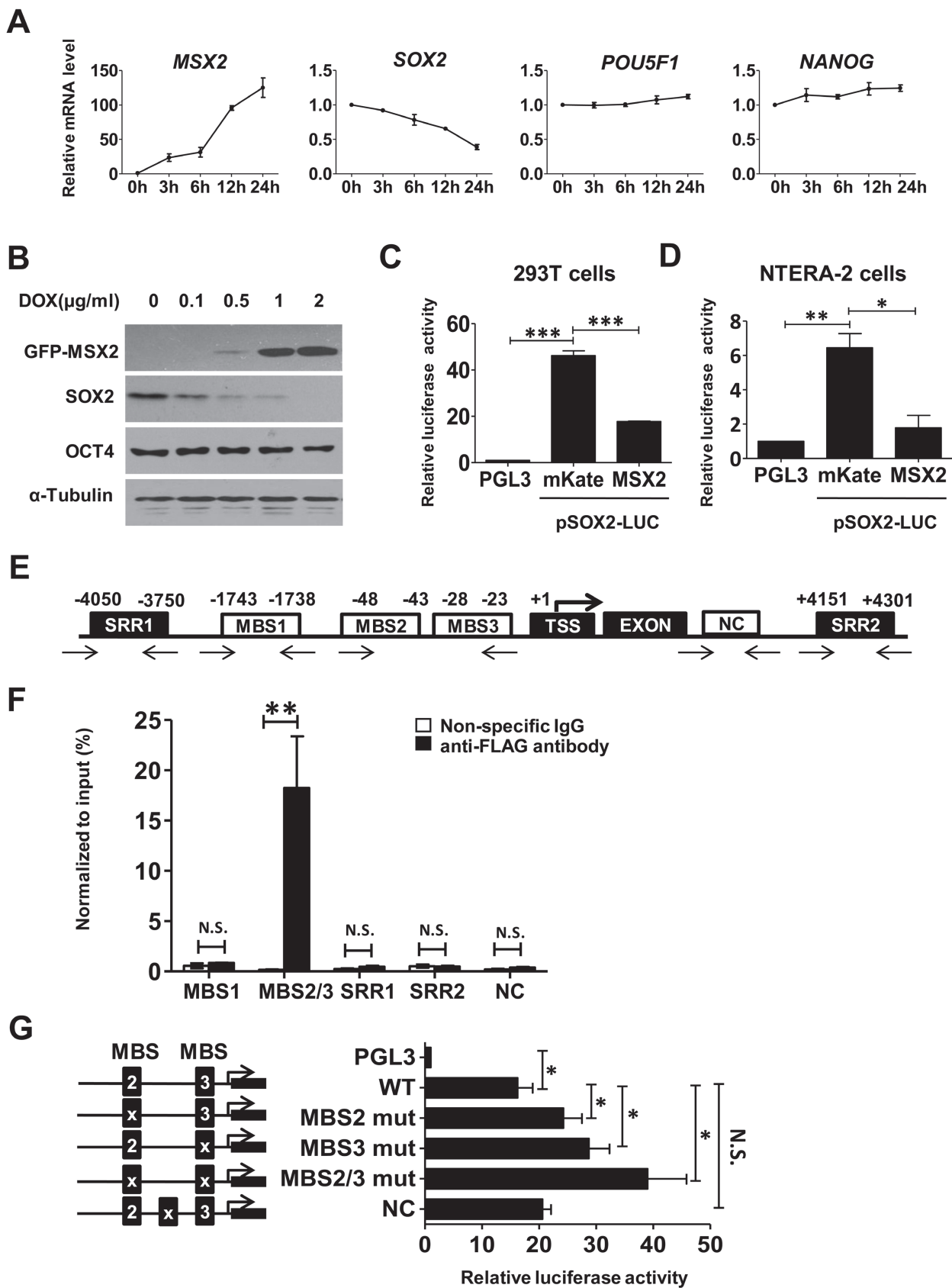
MSX2 suppresses SOX2 via directly binding to SOX2 promoter

How does MSX2 induce mesendoderm differentiation of hESCs? Notably, MSX2 overexpression in hESCs led to the differential expression of a number of cell fate-specifying factors (Figure 1B). Among them, the decrease in SOX2 expression appeared rapid and dramatic (Figure 1B; Supplementary information, Figure S1C and S1D). SOX2 is an integral component of the core transcriptional circuitry of pluripotency and also plays an essential role in neuroectoderm specification of hPSCs [52, 53]. Depletion of SOX2 in hPSCs disrupts pluripotency and induces mesendoderm activities, reminiscent of MSX2 overexpression [54, 55]. We thus asked whether SOX2 is a downstream target of MSX2 in mesendoderm differentiation.

We examined SOX2 response in the first 24 h of MSX2 overexpression. 3 h after the addition of DOX, an increase in *MSX2* mRNA level was detected (Figure 4A). Strikingly, a concomitant decrease in *SOX2* mRNA level was also seen at this time point. As *MSX2* level continued to increase at later time points, the decrease in SOX2 level became more drastic (Figure 4A). Furthermore, the negative regulation of SOX2 by MSX2 was seen in a DOX dose-dependent manner (Figure 4B). In contrast, the levels of POU5F1/OCT4 and NANOG were largely unaltered or slightly increased (Figure 4A and 4B). Thus, SOX2 might serve as a direct target of MSX2 in hPSCs during mesendoderm differentiation.

To support this hypothesis, we isolated *SOX2* 5' flank-

Figure 4 MSX2 suppresses SOX2 expression by direct binding to its promoter. **(A)** Time-course analysis of *MSX2*, *SOX2*, *POU5F1* and *NANOG* mRNA levels by real-time PCR in H1 hESCs with MSX2 overexpression induced by DOX (0–24 h). All values are normalized to the level (= 1) of mRNA in the cells cultured in mTeSR1 before DOX treatment (0 h). Results are shown as means \pm SEM ($n = 3$). **(B)** Western blotting analysis of GFP-MSX2, SOX2 and OCT4 proteins in H1 hESCs 48 h after addition of DOX with a dose gradient. α -Tubulin was used as a loading control. A typical experiment from three separate experiments is shown. **(C, D)** Relative luciferase activity in 293T cells **(C)** and NTERA-2 cells **(D)** transfected with pGL3 basic luciferase construct or pGL3 construct containing SOX2 promoter (pSOX2-LUC). Expression vectors for MSX2 or mKate were co-transfected in cells expressing pSOX2-LUC. All values are normalized to the level (= 1) of the luciferase activity in cells transfected with pGL3 basic vector. Results are shown as means \pm SEM ($n = 3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. **(E)** Schematics of the locations of two SOX2 enhancers (SRR1 and SRR2) and three potential MSX2-binding sites (MBS1–3) within SOX2 genomic locus. The transcriptional start site (TSS +1) is shown. Arrows represent primers designed for the two SOX2 enhancers and three predicted MSX2-binding sites. Primers for a non-specific element are designed as a negative control (NC). **(F)** ChIP-qPCR analysis of MSX2 binding to the enhancers (SRR1 and SRR2), three potential MSX2-binding sites (MBS1, MBS2 and MBS3) and a non-specific binding element (NC) in SOX2 promoter in H1 hESCs transfected with GFP-FLAG-MSX2 and treated with DOX (2 μ g/ml) for 48 h. Non-specific IgG was used as isotype control. All values are normalized to that of their corresponding input samples. Results are shown as means \pm SEM ($n = 3$). ** $P < 0.01$; NS, not significant. **(G)** Relative luciferase activity in 293T cells transfected with pGL3 vector, or WT SOX2 promoter-luciferase reporter construct, or MSX2-binding site mutated (MBS2 mut, MBS3 mut, MBS2/3 mut) SOX2 promoter-luciferase reporter constructs. These cells were co-transfected with a MSX2 expression vector. A non-specific mutant in SOX2 5' flanking region was used as a negative control (NC). All values were normalized to the level (= 1) of the luciferase activity in cells transfected with pGL3 vector. Results are shown as means \pm SEM ($n = 3$). * $P < 0.05$; NS, not significant. Results from three separate experiments are shown as means \pm SEM (see also Supplementary information, Figure S4).



ing sequences of various lengths (0.5, 1 and 2 kb) and tested their responses to enforced MSX2 expression using luciferase reporter assay. Indeed, all three *SOX2* promoter fragments responded to MSX2 overexpression by decreasing the luciferase activity (Supplementary information, Figure S4E). Because no significant difference was observed among the three fragments, we used the 0.5 kb promoter fragment for subsequent analysis.

We found that MSX2 overexpression significantly reduced activity of *SOX2* promoter both in HEK293T cells (Figure 4C) and in NTERA-2 cells (a human pluripotent embryonal carcinoma cell line; Figure 4D) in a dose-dependent manner (Supplementary information, Figure S4F). In contrast, MSX1 overexpression had minimal effect on *SOX2* promoter activity (Supplementary information, Figure S4G). Bioinformatics analysis identified three potential MSX2 binding sequences (MBS1-3) within the 2 kb *SOX2* promoter region [56] (Figure 4E). In addition to the MBSs, we also included two previously reported *SOX2* transcription enhancer sequences [57] in our analysis (designated as SRR1 and SRR2; Figure 4E). ChIP-qPCR in H1 hESCs revealed that MSX2 bound to MBS2 and MBS3, but not MBS1, SRR1 or SRR2 (Figure 4F).

We next asked whether MSX2 binding to MBS2 and MBS3 was responsible for suppression of *SOX2* transcription. By using site-directed mutagenesis, we generated mutations in MBS2 and MBS3, separately or together (Supplementary information, Figure S4H), and found that MSX2 suppression of *SOX2* promoter activity was substantially attenuated, especially when both MBS2 and MBS3 were mutated (Figure 4G). In keeping with the effect of mutations, deletion of MBS2 and MBS3 also impaired MSX2 suppression of *SOX2* (Supplementary information, Figure S4I). Together, these data indicate that MSX2 suppresses *SOX2* expression by directly binding to and inhibiting *SOX2* promoter.

MSX2 suppression of SOX2 is linked to its mesendoderm-inducing function

We next asked whether MSX2 suppression of *SOX2* is linked to its function in mesendoderm induction. To this end, we created two MSX2 mutants: MSX2-T147A and MSX2-P148H, which can reduce and enhance DNA binding activities, respectively [28] (Figure 5A). Surprisingly, both mutants exhibited normal binding activities to MBS2 and MBS3 in *SOX2* promoter and were still capable of inhibiting *SOX2* promoter activity and suppressing *SOX2* expression (Figure 5B and Supplementary information, Figure S5A). Consistently, overexpression of these mutants induced hESCs to undergo mesendoderm differentiation (Figure 5C and 5D, Supplementary

information, Figure S5A and S5B).

We next generated another MSX2 mutant containing a deletion (Δ 132-148; Figure 5A) within the homeodomain that carries core suppressor function of MSX2 [28, 58]. In contrast to the T147A and P148H mutants, the MSX2- Δ 132-148 mutant failed to bind to MBS2 and MBS3 (Supplementary information, Figure S5C) and could not inhibit *SOX2* promoter activity or induce mesendoderm differentiation of hESCs (Figure 5B-5D). We noticed that the expression level of MSX2- Δ 132-148 was slightly lower than the other two mutants (Supplementary information, Figure S5B). To exclude the possibility that the lack of function was due to low expression, we induced MSX2- Δ 132-148 expression using different concentrations of DOX, which led to a dose-dependent increase of MSX2- Δ 132-148 protein (Supplementary information, Figure S5D and S5E). There was no induction of mesendoderm differentiation even with a high DOX concentration (4 μ g/ml; Figure 5C, 5D and Supplementary information, Figure S5E). In addition, prolonged induction of MSX2- Δ 132-148 expression (for 120 h) also failed to induce mesendoderm differentiation of hESCs, suggesting that the lack of effect was not due to a slower response (Supplementary information, Figure S5F).

Mutual antagonism between SOX2 and MSX2

Having established that MSX2 inhibits *SOX2* expression by directly binding to *SOX2* promoter, we investigated the possibility that *SOX2* might also suppress MSX2 level based on the following correlative evidence. First, *SOX2* is highly expressed in hESCs under self-renewal condition, while MSX2 is almost undetectable. Second, in hESCs induced to undergo neural differentiation, *SOX2* remains at a high level, but MSX2 expression is continuously suppressed.

We first asked whether enforced *SOX2* expression could reduce the level of MSX2 and rescue MSX2-induced hESCs mesendoderm differentiation. We fused *SOX2* to mKate, a red fluorescence protein, and expressed the chimeric gene in cells overexpressing MSX2. Remarkably, enforced expression of *SOX2* nearly completely rescued the effect caused by MSX2 expression: cells with *SOX2* overexpression exhibited colony morphologies highly reminiscent of those of normal hESCs (Figure 6A). The rescue effect appeared specific for *SOX2*: enforced expression of OCT4 failed to reverse MSX2-induced morphological changes (Supplementary information, Figure S6A). At the molecular level, *SOX2* overexpression prevented the upregulation of mesendoderm markers such as T and MIXL1, as determined by real-time PCR and western blotting analyses (Figure 6B and 6C). Although OCT4 overexpression attenuated T

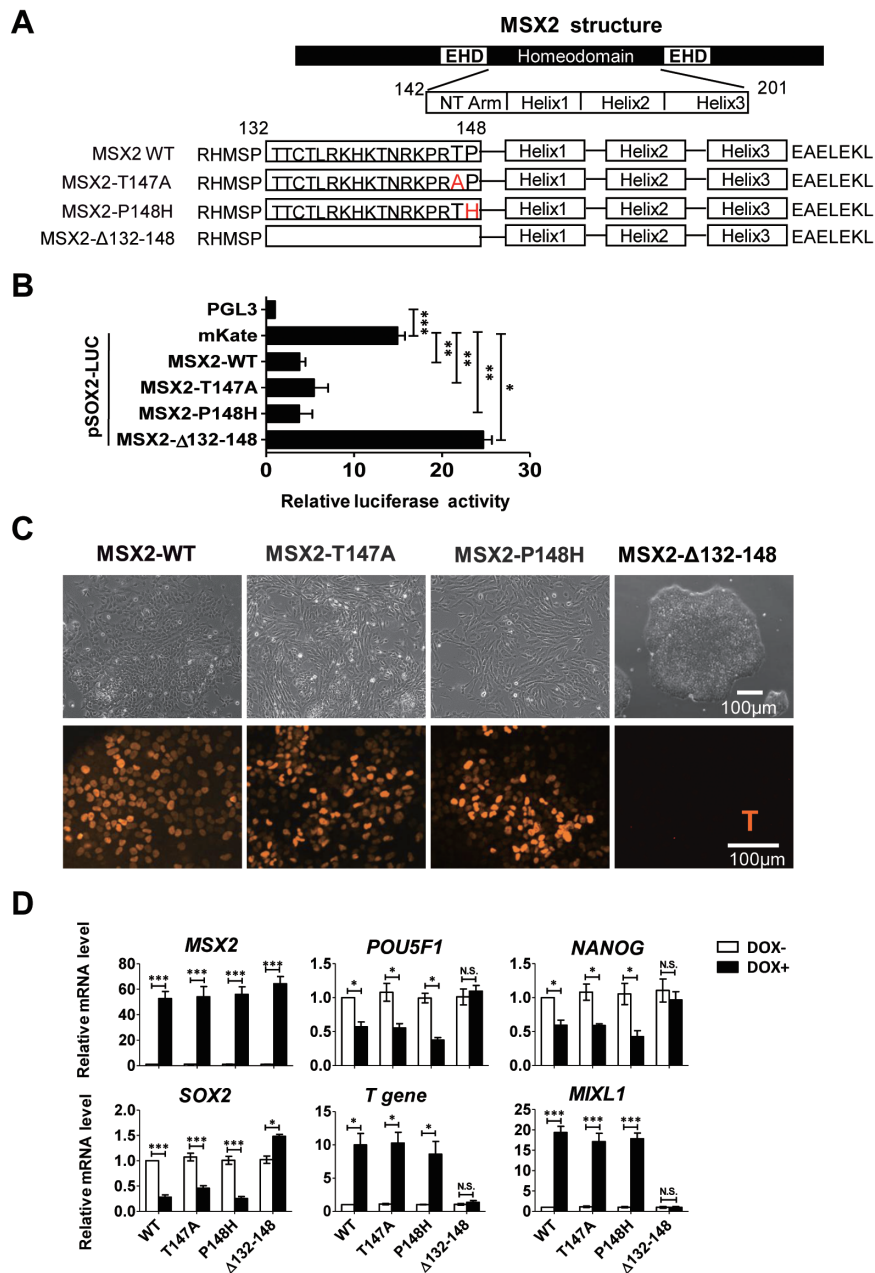


Figure 5 MSX2 suppression of SOX2 is linked to its mesendoderm-inducing function. **(A)** Primary structure of WT MSX2 protein and MSX2 mutants. The homeodomain contains three helices (helix1-3). NT Arm, homeodomain N-terminal arm. **(B)** Relative luciferase activity in 293T cells transfected with pGL3 vector or SOX2 promoter-luciferase reporter construct (pSOX2-LUC). mKate, or WT MSX2, or MSX2 mutant expression vector was co-transfected with pSOX2-LUC. All values are normalized to the level (= 1) of the luciferase activity in cells transfected with pGL3 vector. Results are shown as means \pm SEM ($n = 3$). * $P < 0.05$; ** $P < 0.01$, *** $P < 0.001$. **(C)** Phase contrast (top) and fluorescence (bottom) images of T expression (orange) in H1 hESCs induced to express GFP-MSX2-WT, GFP-MSX2-T147A, GFP-MSX2-P148H or GFP-MSX2- Δ 132-148 upon the addition of DOX (4 μ g/ml, 120 h). **(D)** Real-time PCR analysis of gene expression in H1 hESCs induced to express MSX2-WT or mutants upon the addition of DOX (4 μ g/ml, 120 h). All values are normalized to the level (= 1) of mRNA in H1 cells cultured in mTeSR1 without DOX addition (WT DOX-). Results are shown as means \pm SEM ($n = 3$). * $P < 0.05$; ** $P < 0.01$, *** $P < 0.001$. NS, not significant (see also Supplementary information, Figure S5).

upregulation induced by MSX2, it failed to rescue expression of the pluripotency markers SOX2, endogenous POU5F1/OCT4 and NANOG, which was even further downregulated (Supplementary information, Figure S6B). Strikingly, SOX2 overexpression also significantly impaired enforced expression of MSX2, as shown by a

substantial reduction in the level of GFP-MSX2 protein (Figure 6C). Interestingly, despite the reduced level of GFP-MSX2, the *MSX2* mRNA level remained largely unaltered (Figure 6B), implying that SOX2 regulation of MSX2 occurs post-transcriptionally. Indeed, it has been reported that the presence of MSX2-P148H mu-

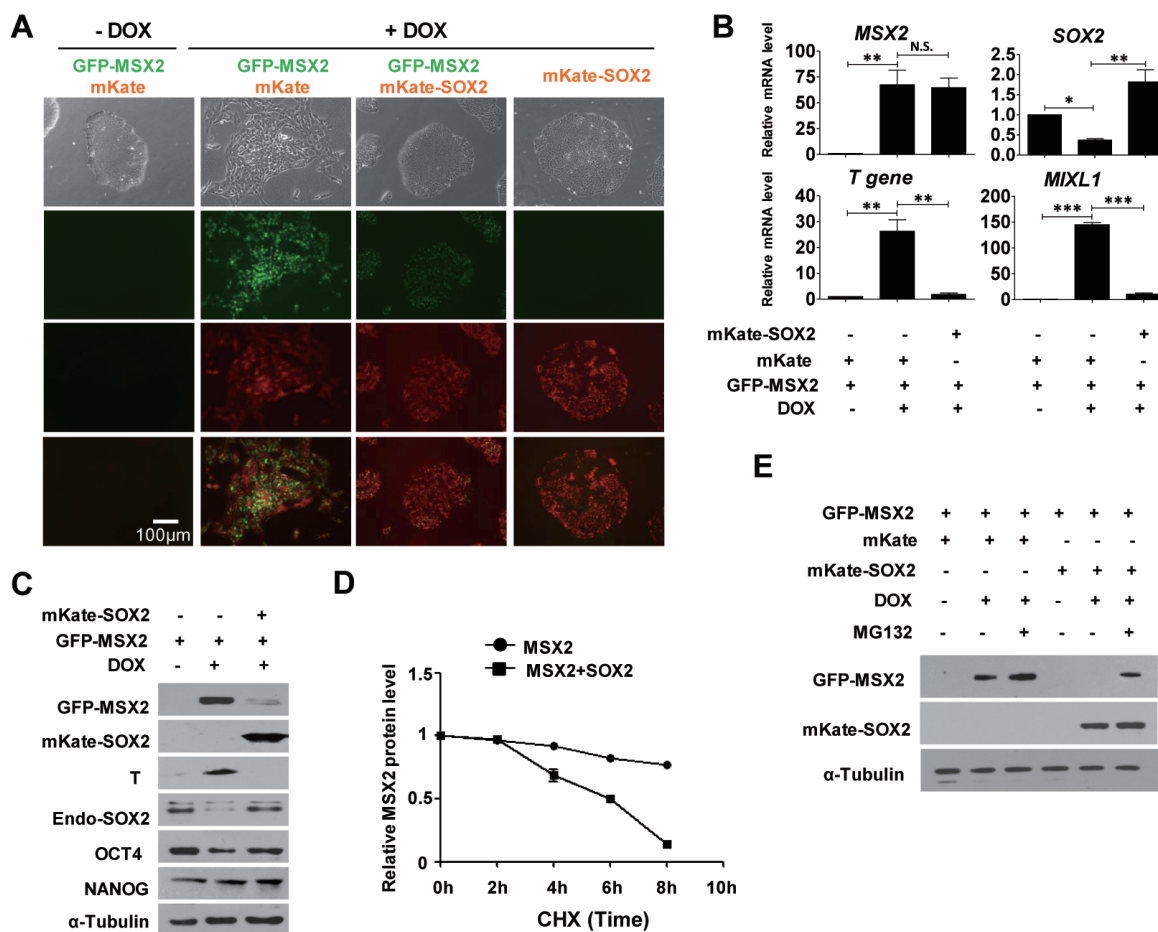


Figure 6 SOX2 rescues MSX2 mesendoderm induction by promoting its degradation. **(A)** Phase contrast and fluorescence images of H1 hESCs with or without the induction of GFP-MSX2 expression by DOX. Cells were also infected with ptight-mKate or ptight-mKate-SOX2 inducible expression vectors. Fluorescence images of GFP and mKate are merged (bottom). Scale bar, 100 µm. **(B)** Real-time PCR analysis of *MSX2*, *SOX2*, *T* and *MIXL1* mRNA levels in H1 hESCs with or without the induction of GFP-MSX2 expression by DOX (72 h). Cells were also infected with ptight-mKate or ptight-mKate-SOX2 inducible expression vectors. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant. All values are normalized to the level (= 1) of mRNA in H1 cells cultured in mTeSR1 without DOX addition. Results are shown as means \pm SEM ($n = 3$). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant. **(C)** Western blotting analysis of various proteins in H1 cells under condition similar to **(A)** and **(B)**. α -Tubulin was used as a loading control. A typical experiment from three separate experiments is shown. **(D)** Relative level of MSX2 protein in H1 hESCs induced to express GFP-MSX2, with or without mKate-SOX2. Cells were analyzed at different time points after CHX treatment (100 µg/ml). All values are normalized to the level (= 1) of protein in H1 cells cultured in mTeSR1 without DOX addition (also see the Supplementary information, Figure S6F). **(E)** Western blotting analysis of mKate-SOX2 and GFP-MSX2 protein in H1 cells induced to express GFP-MSX2 and mKate or express GFP-MSX2 and mKate-SOX2 in the presence or absence of MG132 (5 µM, 4 h). α -Tubulin was used as a loading control. A typical experiment from three separate experiments is shown (see also Supplementary information, Figure S6).

tant causes higher susceptibility to ubiquitin-dependent degradation of wild-type MSX2 protein [59], implying protein degradation as a potential regulatory mechanism of MSX2 expression.

We explored how SOX2 overexpression might lead to decreased MSX2 protein level. First, in both HEK293T cells and hESCs, expression of SOX2 decreased the level of MSX2 in a dose- and time-dependent manner (Supplementary information, Figure S6C and S6D). In contrast, SOX2 overexpression failed to decrease the level of the MSX2- Δ 132-148 mutant protein (Supplementary information, Figure S6E). Next, to examine whether SOX2 promotes the degradation of MSX2 protein, we assessed the half-life of MSX2 in the presence of protein synthesis inhibitor cycloheximide. As shown in Figure 6D and Supplementary information, Figure S6F, the half-life of MSX2 was slightly reduced in control cells without SOX2 overexpression, and was substantially reduced with SOX2 overexpression. Furthermore, because intracellular protein turnover is primarily mediated by proteasomal degradation, we treated H1 hESCs with MG132, a potent inhibitor of the proteasome. MG132 slightly increased the level of MSX2 protein in the absence of SOX2 overexpression, but almost completely rescued the reduction of MSX2 protein caused by SOX2 overexpression (Figure 6E). Thus, SOX2 promotes MSX2 protein degradation via the proteasome pathway.

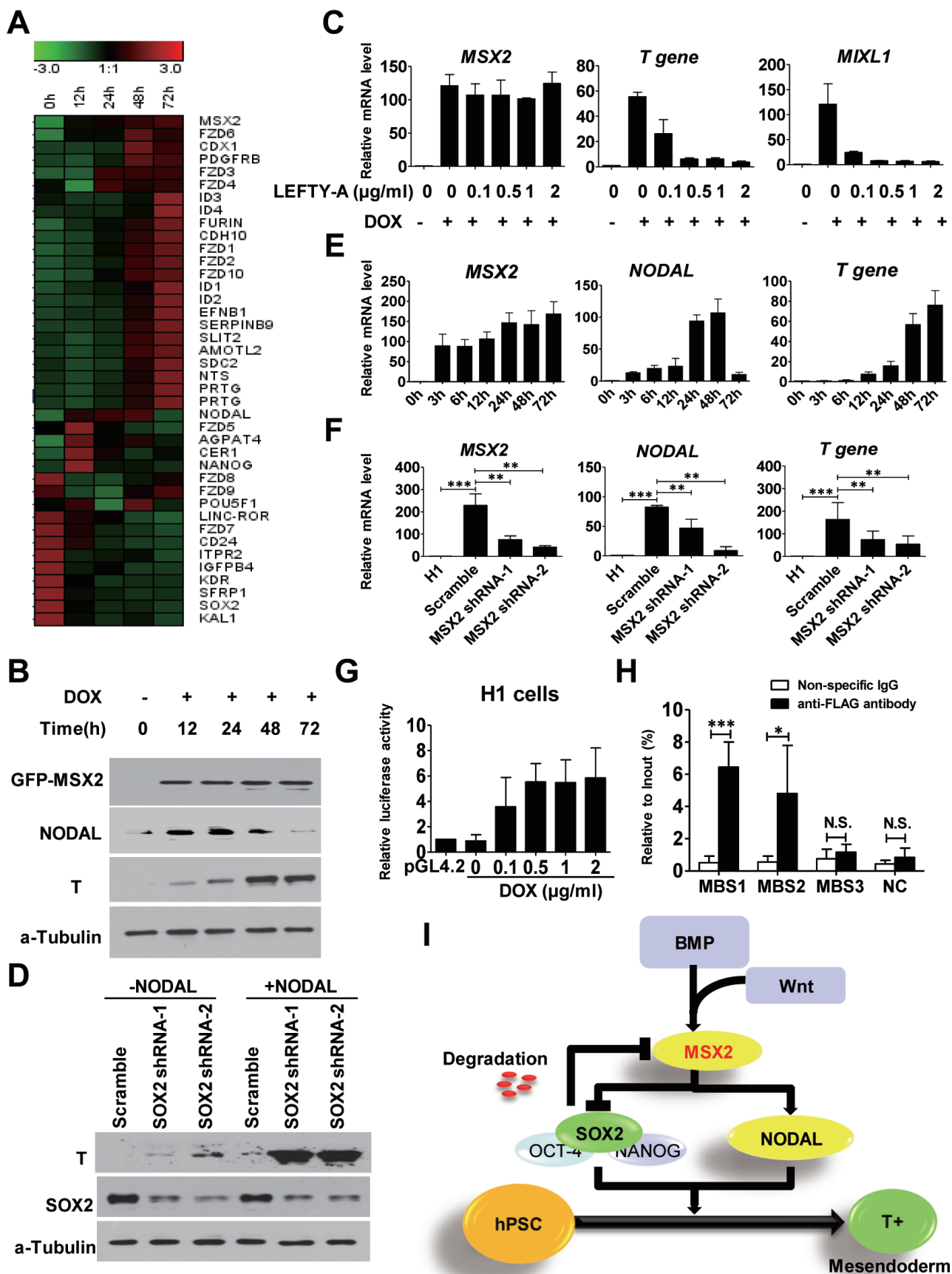
Taken together, our results reveal a mutual antagonism between MSX2 and SOX2 in hPSCs. SOX2 does not merely act downstream of MSX2 but can also destabilize MSX2 protein. This mutual antagonism likely plays a significant role in fate decisions of hPSCs.

Nodal signaling is an essential downstream effector of MSX2

We have showed that MSX2 overexpression leads to rapid downregulation of SOX2 in hESCs (Figures 1B, 4A and 4B). If SOX2 downregulation is solely responsible for MSX2 induction of mesendoderm differentiation, SOX2 depletion in hESCs should mimic the effects of MSX2 enforced expression. Interestingly, we found that although SOX2 depletion disrupted pluripotency and induced mesendoderm activities in hESCs, the upregulation of mesendoderm markers upon SOX2 depletion was much less rapid and drastic than enforced MSX2 expression (Supplementary information, Figure S7A). These results are consistent with an earlier study [54]. We therefore inferred that additional pathway(s) might mediate the function of MSX2 in mesendoderm induction.

To discover additional MSX2 downstream effectors, we conducted whole genome RNA sequencing (RNA-seq) in hESCs after induction of MSX2 expression. We focused our analysis on early time points (0, 12, 24, 48 and 72 h) after the addition of DOX to reveal potential im-

Figure 7 Nodal signaling is an essential downstream effector of MSX2. **(A)** Heat maps of selected genes from RNA-seq analysis for H1 cells induced to express MSX2 by DOX (2 μ g/ml; 0 h to 72 h). Fold of changes in \log_2 scale, is normalized to the level of gene expression in H1 hESCs cultured in mTeSR1 medium without DOX (denoted as 0 h). $-3 \leq \log_2 X \leq 3$. **(B)** Western blotting analysis of GFP-MSX2, NODAL and T protein expression in 72 h in H1 cells induced to express GFP-MSX2 by DOX (2 μ g/ml). α -Tubulin was used as a loading control. A typical experiment from three separate experiments is shown. **(C)** Real-time PCR analysis of *MSX2*, *T* and *MIXL1* expression in H1 cells expressing GFP-MSX2 after DOX induction. Cells were also treated with increasing concentrations of LEFTY-A. All values are normalized to the level (= 1) of mRNA in H1 cells cultured in mTeSR1 without DOX and LEFTY-A. Results from three separate experiments are shown as mean \pm SEM. **(D)** Western blotting analysis of T and SOX2 in H1 hESCs depleted of SOX2 (with shRNA-1 or shRNA-2) with or without NODAL (200 ng/ml). Cells treated with a scramble shRNA (Scramble) served as negative control; α -Tubulin was used as a loading control. A typical experiment from three separate experiments is shown. **(E)** Time-course analysis of *MSX2*, *NODAL* and *T* gene expression by real-time PCR in H1 cells treated with BMP4 (20 ng/ml) combined with Wnt3a (20 ng/ml). All values are normalized to the level (= 1) of mRNA in H1 cells cultured in mTeSR1 before treatment (0 h). Results are shown as means \pm SEM ($n = 3$). **(F)** Real-time PCR analysis of *MSX2*, *NODAL* and *T* gene expression in H1 hESCs depleted of MSX2 depletion (with shRNA-1 or shRNA-2) or infected with a scramble shRNA (Scramble) before (0 h) and 48 h after treatment with BMP4 (20 ng/ml) combined with Wnt3a (20 ng/ml). All values are normalized to the level (= 1) of mRNA in the cells infected with scramble shRNA and cultured in mTeSR1 without treatment. Results are shown as means \pm SEM ($n = 3$). ** $P < 0.01$; *** $P < 0.001$. **(G)** Relative luciferase activity in GFP-MSX2 H1 cells transfected with PGL4.2 basic luciferase construct or pGL4.2 construct containing *NODAL* promoter (pNODAL-LUC). Various levels of GFP-MSX2 expression were induced by increasing doses of DOX. All values are normalized to the level (= 1) of the luciferase activity in cells transfected with pGL4.2 empty vector. Results are shown as means \pm SEM ($n = 3$). **(H)** ChIP-qPCR analysis of MSX2 binding to the three potential MSX2-binding sites (MBS1, MBS2 and MBS3) and a non-specific binding element (NC) of *NODAL* promoter in H1 cells overexpressing GFP-FLAG-MSX2 upon DOX treatment (2 μ g/ml) for 48 h. Non-specific IgG was used as isotype control. All values are normalized to that of their corresponding input samples. Results are shown as means \pm SEM ($n = 3$). * $P < 0.05$; ** $P < 0.01$, *** $P < 0.001$. NS, not significant. **(I)** Working model for MSX2 function and mechanism in hESC early differentiation (see also Supplementary information, Figures S7 and S8).



mediate downstream targets of MSX2. Among the genes with significantly altered expression, NODAL, a member of the TGF- β superfamily, was rapidly (within 12 h) and drastically upregulated after MSX2 induction, and stayed at a high level until 48 h (Figure 7A). The Activin/Nodal signaling pathway is essential for early embryonic development and plays a critical role in mesendoderm differentiation of hPSCs [17, 21], leading us to hypothesize that Nodal signaling is involved in MSX2 induction of mesendoderm differentiation.

Indeed, MSX2 overexpression upregulated NODAL at both the mRNA and protein levels in hESCs (Figure 7B, Supplementary information, Figure S7B). Interestingly, the MSX2- Δ 132-148 mutant, which fails to induce mesendoderm differentiation, also failed to induce the upregulation of NODAL (Supplementary information, Figure S7C). To test the role of NODAL in hESC early differentiation, we used Nodal inhibitor LEFTY-A and found it inhibited MSX2-induced upregulation of mesendoderm markers in a dose-dependent manner (Figure 7C). Similarly, SB-431542, a commonly used chemical inhibitor of TGF- β receptors, also prevented MSX2-induced mesendoderm differentiation (Supplementary information, Figure S7D). NODAL addition to hESCs under self-renewal condition was insufficient to disrupt pluripotency or induce differentiation, in keeping with previously documented function of NODAL in supporting pluripotency [8, 17]. However, NODAL addition to hESCs depleted of SOX2 significantly increased the expression of mesendoderm markers over the level of SOX2 depletion alone (Figure 7D, Supplementary information, Figure S7E). Thus, NODAL in conjunction with SOX2 depletion mimics the effect of enforced MSX2 expression in hESCs, suggesting that SOX2 and NODAL are essential downstream effectors of MSX2.

Interestingly, it has been reported that BMP, Wnt and NODAL can regulate each other and form a signaling loop essential for early lineage specification of mouse and human embryonic stem cells [60-62]. In this study we have shown that MSX2 acts as a direct target gene of BMP4, while Wnt signaling activates MSX2 expression synergistically with BMP. We therefore asked whether MSX2 is involved in BMP and Wnt regulation of NODAL. Indeed, BMP4 and Wnt3a treatment elevated the expression of MSX2, NODAL and T (Figure 7E). Depletion of MSX2 significantly repressed NODAL and T upregulation (Figure 7F, Supplementary information, Figure S8A). Our data suggest that NODAL is a downstream target of MSX2 through BMP and Wnt.

To further explore the mechanism by which MSX2 activates NODAL expression, we isolated 5 kb *NODAL* 5' flanking sequences and tested its responses to MSX2

overexpression using reporter assay. MSX2 overexpression significantly enhanced the *NODAL* promoter activity in both HEK293T cells and hESCs in a dose-dependent manner (Figure 7G and Supplementary information, Figure S8B). Bioinformatics analysis predicted three potential MSX2-binding sequences (MBS1-3) within the 5 kb region of *NODAL* promoter (Supplementary information, Figure S8C). ChIP-qPCR analysis in H1 hESCs revealed that MSX2 directly bound to MBS1 and MBS2 of *NODAL* promoter (Figure 7H). Consistent with its lack of function, MSX2- Δ 132-148 mutant did not have binding activity to the *NODAL* promoter (Supplementary information, Figure S8D). These data suggest that MSX2 induces NODAL expression by directly binding to and activating the *NODAL* promoter.

Discussion

This study identifies MSX2 as a master mediator of hPSCs' differentiation to mesendoderm. MSX2 is both necessary and sufficient for mesendoderm differentiation in hPSCs. MSX2 is a direct target of BMP signaling in hPSCs, and it is also synergistically activated by Wnt signals via LEF1 during mesendoderm differentiation. Mechanistically, MSX2 binds to *SOX2* promoter and inhibits *SOX2* transcription, thus destabilizing the pluripotency circuitry. Interestingly, *SOX2* can antagonize the function of *SOX2* by promoting MSX2 degradation. Furthermore, MSX2 induction of mesendoderm differentiation not only requires *SOX2* suppression but also involves activation of Nodal signaling. Our results provide conclusive evidence that MSX2 is a key determinant of hPSC mesendoderm differentiation (Figure 7I). Notably, our results differ from earlier reports that MSX2 is dispensable in early embryonic development [26, 29]. The lack of phenotype in MSX2 deletion is not due to functional redundancy of MSX1 and MSX2, because MSX1/MSX2 double knockout mice form embryonic germ layers normally [45, 63]. The discrepancy between the human and the mouse studies is likely due to the distinct function of MSX2 in the two species. Similar species-specific functions of cell fate-specifying factors have been reported previously. For example, Zhang *et al.* [64] demonstrated that PAX6 is a key determinant of neuroectoderm cell fate in hPSCs but is not involved in mouse neuroectoderm specification.

Among various developmental pathways, the BMP pathway has profound roles in directed differentiation of hPSCs to the early embryonic lineages such as mesendoderm [16, 23]. Despite these critical functions, the direct target genes of BMP signaling that mediate mesendoderm differentiation remain elusive in hPSCs. In this study,

we show that expression of *MSX2* is induced within 3 h after the addition of *BMP4*, much more early than *T* and *CDX2*, two previously reported downstream effectors of *BMP* signaling in hESCs [48]. We show that *Smad1* can bind to the phylogenetically conserved *BMP*-responsive region within the *MSX2* promoter in hESCs. Thus, *MSX2* can directly respond to *BMP* signaling in hPSCs to mediate *BMP* functions in mesendoderm induction. Furthermore, *MSX2* not only responds more quickly to *BMP4* stimulation than *T*, but also functions upstream of *T* during mesendoderm induction of hESCs — overexpression of *MSX2* induces *T* expression, whereas *MSX2* depletion prevents *T* induction. These results together indicate that *MSX2* is an early and direct target of *BMP* in hPSC mesendoderm differentiation. In addition to *BMP*, *MSX2* expression can be further enhanced by *Wnt* signals via *LEF1*, likely by cooperating with phosphorylated *Smads* as previously reported in mouse ESCs [51]. This latter result suggests that *MSX2* acts as a central signaling molecule integrating multiple extrinsic signals during hPSC mesendoderm induction.

How does *MSX2* mediate mesendoderm differentiation? We provide evidence for *SOX2* as a direct downstream target of *MSX2* in hESCs during mesendoderm induction. *MSX2* overexpression rapidly reduces the level of *SOX2* mRNA and protein. At the mechanistic level, *MSX2* directly binds to *MBS2* and *MBS3* within the *SOX2* promoter, leading to inhibition of *SOX2* transcription. This transcriptional repressor function of *MSX2* is crucial for hESC mesendoderm differentiation: deletion mutant (*MSX2*- Δ 132-148), which lacks suppressor activity, also fails to induce mesendoderm differentiation. The link between *MSX2* and *SOX2* allows us to postulate a mechanism by which *BMP* signaling can destabilize the pluripotency state. We envision that *MSX2*, once activated by mesendoderm-inducing cues such as *BMPs*, represses *SOX2* transcription, thus leading to a perturbation of the pluripotency circuitry and exit of hPSCs from the pluripotent state. In addition, the continuous suppression of *SOX2* by *MSX2* is likely needed for the inhibition of neuroectoderm activity during differentiation, thus allowing hPSC differentiation to be restricted to mesendoderm. In support of this notion, depletion of *SOX2* disrupts pluripotency and induces mesendoderm differentiation of hPSCs.

Interestingly, although *SOX2* depletion induces mesendoderm in hPSCs, the upregulation of mesendoderm markers is much less rapid and drastic than by enforced *MSX2* expression. Thus, suppression of *SOX2* expression by *MSX2* cannot fully recapitulate the ability of *MSX2* to induce mesendoderm differentiation in hPSCs. By applying a functional genomics approach, we identi-

fy *NODAL* as a novel direct downstream target gene of *MSX2* essential for mesendoderm differentiation. In this scenario, *BMP* and *Wnt* trigger *MSX2* expression, which in turn induces *NODAL* expression via direct binding and activation of its promoter activity. Multiple studies have shown that *BMP*, *Wnt* and *NODAL* can regulate each other and form a signaling network essential for early lineage specification of mouse and human embryonic stem cells [21, 61]. However, how *BMP* and *Wnt* regulate *NODAL* expression during different development stages is still being defined. Here, we present evidence that *MSX2* mediates *BMP* and *Wnt* regulation of *NODAL* function in hESCs during early lineage specification. Although *Nodal* addition alone has little effect on hPSCs under self-renewal condition, it drastically enhances the effect of *SOX2* depletion on hESC mesendoderm differentiation. The differential effects of *NODAL* in cells with or without *SOX2* depletion are reminiscent of previous findings that *Nodal/Activin* signaling supports pluripotency in the presence of *bFGF* but induces mesendoderm differentiation when *bFGF* is removed [17, 20].

Importantly, *SOX2* does not simply lie downstream of *MSX2* but can also inhibit the function of *MSX2* by reducing the level of *MSX2* protein. We further show that *SOX2* promotes *MSX2* protein degradation via a *MG132*-sensitive proteasomal pathway. Interestingly, *SOX2* fails to decrease the level of *MSX2*- Δ 132-148 mutant protein suggesting the deleted amino acids are not only required for the core suppressor function of *MSX2* but also mediate *MSX2* degradation. We envision that the mutual antagonism between the two key lineage-specifying factors plays a significant role in fate determination in hPSCs. *MSX2* suppression of *SOX2* is essential for mesendoderm differentiation. Conversely, *SOX2* inhibition of *MSX2* contributes to the surveillance of pluripotency, and promotes neuroectoderm differentiation of hPSCs when *MSX2* expression is down-regulated. Our findings add to emerging evidence that mutual repression between cell-fate regulators represents a key mechanism for governing fate decisions of stem cells. For example, mutual antagonism between *OCT4* and *CDX2* determines the segregation of the inner cell mass (ICM) and trophectoderm [65], while *NANOG* and *GATA6* antagonize each other to control ICM cell fate transition to epiblast or the primitive endoderm [66]. Notably, mutual repression between *OCT4* and *CDX2* occurs at the transcriptional level, while *MSX2* and *SOX2* cross-inhibition exhibits a more complex pattern, occurring at the transcriptional and the protein level.

Materials and Methods

Cell culture

H1 and H9 hESCs (passage 30-60) were purchased from the WiCell Research Institute (Madison, WI, USA) and maintained and expanded in mTeSR1 medium (StemCell Technologies), or E8 medium (Gibco), or MEF-CM changed daily. For differentiation, cells were grown in custom mTeSR1 medium (StemCell Technologies) with or without selected factors as described in the figure legends. NTERA-2 cells were obtained from American Type Culture Collection (ATCC) and cultured according to ATCC's recommendation. The culture medium consists of 10% fetal bovine serum (ATCC). HEK293T cells (ATCC) were cultured according to ATCC recommended protocol. For additional details, see Supplementary information, Data S1.

Lentivirus production and hESC infection

Lentiviruses for gene knockdown or overexpression were packaged using Viralpower Lentivirus Packaging System (Invitrogen) according to the manufacturer's instruction. To infect hESCs, lentiviruses were mixed with mTeSR1 medium, and the mixture was incubated with H1 or H9 cells ($\sim 1 \times 10^5$ cells) for 24 h. A multiplicity of infection of 100 was used. To improve the efficiency of infection, 3 $\mu\text{g/ml}$ polybrene (Sigma) was added to the infection medium. To establish hESC stable line with inducible gene expression, Lenti-X™ Tet-On Advanced Inducible Expression System (Clontech) was used following the manufacturer's instruction. For additional details, see Supplementary information, Data S1.

Teratoma formation in SCID mice

All mice used in this study were maintained under specific pathogen-free conditions and all procedures have been approved by the Peking Union Medical College Institutional Animal Care and Use Committee. H1 hESCs containing GFP-MSX2 or GFP inducible expression vectors were harvested and subcutaneously injected into the hind leg region of SCID mice. For additional details, see Supplementary information, Data S1.

Western blotting

For protein analysis, 5×10^6 cells were lysed directly in 200 μl laemmli sample buffer (BioRad). Dilutions for various antibodies were described in Supplementary information, Table S4. The blots were developed using Super-Signal West Pico Chemiluminescent Substrate (Pierce), and signals were quantified with Image J. For additional details, see Supplementary information, Data S1.

Real-time PCR

Total RNA was extracted with Trizol (Invitrogen) following the protocol provided by the manufacturer. Samples were treated with DNase, and the RNA was purified using RNeasy mini kit (Qiagen). cDNA was produced using a reverse transcription system from Promega. All real-time PCR assays were performed with QuantiTect SYBR Green PCR kit (Qiagen) using an ABI PRISM 7900 (Applied Biosystems) following manufacturer's instructions. The primers used are as described in Supplementary information, Table S5.

Reporter assay

Human *MSX2* 2 kb fragment (-4403 to -2335) containing BM-Pre, human *SOX2* gene fragment containing -2 000 to +231, and human *NODAL* gene fragment containing -4862 to +186 were am-

plified from genomic DNA of H1 hESCs and cloned into pGL4.2 or pGL3 basic vector (Promega).

Details for the construction of pBMPPre-LUC, pSOX2-LUC, deletion mutants, *MSX2* binding-site mutants, and pNODAL-LUC are available in Supplementary information, Data S1. Cell transfection and cell lysate preparation were performed as described in Supplementary information, Data S1. Firefly luciferase and Renilla luciferase activities were analyzed using a Synergy H4 hybrid Reader (BioTek) according to manufacturer's instruction (Promega). For additional details, see Supplementary information, Data S1.

Chromatin immunoprecipitation (ChIP)

ChIP and quantitative PCR were performed following standard methods. For *MSX2* binding sites on *SOX2* promoter or *NODAL* promoter, H1 hESCs overexpressing GFP-FLAG-*MSX2* or GFP-FLAG-*MSX2*- $\Delta 132$ -148 were crosslinked with 1% formaldehyde at 37 °C for 10 min. For SMAD binding sites on *MSX2*, H1 hESCs were cultured in custom mTeSR1 medium with 20 ng/ml BMP4 for 6 h prior to fixation. Antibodies and primers are detailed in Supplementary information, Table S7. Values obtained from immunoprecipitated samples are normalized to that of their corresponding input samples. Data are representative of three separate experiments, and error bars indicate mean \pm SEM.

Statistical analysis

All data are presented as mean \pm SEM. Student's *t*-test was used to compare two experimental groups, assuming unequal variances. Differences are considered significant when $P < 0.05$.

Accession number

The RNA-Seq data have been deposited in NCBI (accession number: SRP055541).

Acknowledgements

This work was supported by the National Basic Research Program of China (2012CB966403 and 2015CB964902 to JZ), the National Natural Science Foundation of China (31171431 to JZ, 31301206 to YB and 81370600 to DL) and Tianjin Natural Science Foundation (11JCYBJC27300 to JZ). We thank Drs. Tao Cheng, Dangsheng Li and Lihong Shi for their suggestions during manuscript preparation. We thank members of the Zhou lab for their insightful discussion during the course of this work. We thank Drs Xiaoping Zhang and Baoyang Hu for providing plasmids for this study.

References

- 1 Thomson JA, Itskovitz-Eldor J, Shapiro SS, *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* 1998; **282**:1145-1147.
- 2 Yu J, Vodyanik MA, Smuga-Otto K, *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; **318**:1917-1920.
- 3 Rossant J. Stem cells and early lineage development. *Cell* 2008; **132**:527-531.
- 4 Yamanaka S. A fresh look at iPS cells. *Cell* 2009; **137**:13-17.
- 5 Boyer LA, Lee TI, Cole MF, *et al.* Core transcriptional reg-

- ulatory circuitry in human embryonic stem cells. *Cell* 2005; **122**:947-956.
- 6 Dalton S. Signaling networks in human pluripotent stem cells. *Curr Opin Cell Biol* 2013; **25**:241-246.
 - 7 Vallier L, Mendjan S, Brown S, *et al.* Activin/Nodal signaling maintains pluripotency by controlling Nanog expression. *Development* 2009; **136**:1339-1349.
 - 8 James D, Levine AJ, Besser D, Hemmati-Brivanlou A. TGF-beta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* 2005; **132**:1273-1282.
 - 9 Armstrong L, Hughes O, Yung S, *et al.* The role of PI3K/AKT, MAPK/ERK and NFkappabeta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. *Hum Mol Genet* 2006; **15**:1894-1913.
 - 10 Zhou J, Su P, Wang L, *et al.* mTOR supports long-term self-renewal and suppresses mesoderm and endoderm activities of human embryonic stem cells. *Proc Natl Acad Sci USA* 2009; **106**:7840-7845.
 - 11 Singh AM, Reynolds D, Cliff T, *et al.* Signaling network crosstalk in human pluripotent cells: a Smad2/3-regulated switch that controls the balance between self-renewal and differentiation. *Cell Stem Cell* 2012; **10**:312-326.
 - 12 Lian X, Zhang J, Zhu K, Kamp TJ, Palecek SP. Insulin inhibits cardiac mesoderm, not mesendoderm, formation during cardiac differentiation of human pluripotent stem cells and modulation of canonical Wnt signaling can rescue this inhibition. *Stem Cell* 2013; **31**:447-457.
 - 13 Woll PS, Morris JK, Painschab MS, *et al.* Wnt signaling promotes hemoendothelial cell development from human embryonic stem cells. *Blood* 2008; **111**:122-131.
 - 14 Vijayaragavan K, Szabo E, Bosse M, Ramos-Mejia V, Moon RT, Bhatia M. Noncanonical Wnt signaling orchestrates early developmental events toward hematopoietic cell fate from human embryonic stem cells. *Cell Stem Cell* 2009; **4**:248-262.
 - 15 Xu RH, Chen X, Li DS, *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol* 2002; **20**:1261-1264.
 - 16 Zhang P, Li J, Tan Z, *et al.* Short-term BMP-4 treatment initiates mesoderm induction in human embryonic stem cells. *Blood* 2008; **111**:1933-1941.
 - 17 Vallier L, Alexander M, Pedersen RA. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* 2005; **118**:4495-4509.
 - 18 Beattie GM, Lopez AD, Bucay N, *et al.* Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. *Stem Cell* 2005; **23**:489-495.
 - 19 Ramos-Mejia V, Melen GJ, Sanchez L, *et al.* Nodal/Activin signaling predicts human pluripotent stem cell lines prone to differentiate toward the hematopoietic lineage. *Mol Ther* 2010; **18**:2173-2181.
 - 20 Brown S, Teo A, Pauklin S, *et al.* Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cell* 2011; **29**:1176-1185.
 - 21 Chng Z, Teo A, Pedersen RA, Vallier L. SIP1 mediates cell-fate decisions between neuroectoderm and mesendoderm in human pluripotent stem cells. *Cell Stem Cell* 2010; **6**:59-70.
 - 22 Zhou J, Su P, Li D, Tsang S, Duan E, Wang F. High-efficiency induction of neural conversion in human ESCs and human induced pluripotent stem cells with a single chemical inhibitor of transforming growth factor beta superfamily receptors. *Stem Cell* 2010; **28**:1741-1750.
 - 23 Richter A, Valdimarsdottir L, Hrafnskeldottir HE, *et al.* BMP4 promotes EMT and mesodermal commitment in human embryonic stem cells via SLUG and MSX2. *Stem Cell* 2014; **32**:636-648.
 - 24 Bell JR, Noveen A, Liu YH, *et al.* Genomic structure, chromosomal location, and evolution of the mouse Hox 8 gene. *Genomics* 1993; **17**:800.
 - 25 Davidson D. The function and evolution of Msx genes: pointers and paradoxes. *Trends Genet* 1995; **11**:405-411.
 - 26 MacKenzie A, Ferguson MW, Sharpe PT. Expression patterns of the homeobox gene, Hox-8, in the mouse embryo suggest a role in specifying tooth initiation and shape. *Development* 1992; **115**:403-420.
 - 27 Li Y, Liu J, Hudson M, Kim S, Hatch NE. FGF2 promotes Msx2 stimulated PC-1 expression via Frs2/MAPK signaling. *J Cell Biochem* 2010; **111**:1346-1358.
 - 28 Newberry EP, Latifi T, Battaile JT, Towler DA. Structure-function analysis of Msx2-mediated transcriptional suppression. *Biochemistry* 1997; **36**:10451-10462.
 - 29 Satokata I, Ma L, Ohshima H, *et al.* Msx2 deficiency in mice causes pleiotropic defects in bone growth and ectodermal organ formation. *Nat Genet* 2000; **24**:391-395.
 - 30 Ma L, Golden S, Wu L, Maxson R. The molecular basis of Boston-type craniosynostosis: the Pro148-->His mutation in the N-terminal arm of the MSX2 homeodomain stabilizes DNA binding without altering nucleotide sequence preferences. *Hum Mol Genet* 1996; **5**:1915-1920.
 - 31 Garcia-Minaur S, Mavrogiannis LA, Rannan-Eliya SV, *et al.* Parietal foramina with cleidocranial dysplasia is caused by mutation in MSX2. *Eur J Hum Genet* 2003; **11**:892-895.
 - 32 Jabs EW, Muller U, Li X, *et al.* A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. *Cell* 1993; **75**:443-450.
 - 33 Wuyts W, Reardon W, Preis S, *et al.* Identification of mutations in the MSX2 homeobox gene in families affected with foramina parietalia permagna. *Hum Mol Genet* 2000; **9**:1251-1255.
 - 34 Quinn LM, Latham SE, Kalionis B. The homeobox genes MSX2 and MOX2 are candidates for regulating epithelial-mesenchymal cell interactions in the human placenta. *Placenta* 2000; **21 Suppl A**:S50-54.
 - 35 di Bari MG, Ginsburg E, Plant J, Strizzi L, Salomon DS, Vonderhaar BK. Msx2 induces epithelial-mesenchymal transition in mouse mammary epithelial cells through upregulation of Cripto-1. *J Cell Physiol* 2009; **219**:659-666.
 - 36 Friedmann Y, Daniel CW. Regulated expression of homeobox genes Msx-1 and Msx-2 in mouse mammary gland development suggests a role in hormone action and epithelial-stromal interactions. *Dev Biol* 1996; **177**:347-355.
 - 37 Xia X, Ayala M, Thiede BR, Zhang SC. *In vitro*- and *in vivo*-induced transgene expression in human embryonic stem cells and derivatives. *Stem Cells* 2008; **26**:525-533.
 - 38 Suzuki A, Ueno N, Hemmati-Brivanlou A. *Xenopus msx1* mediates epidermal induction and neural inhibition by BMP4.

- Development* 1997; **124**:3037-3044.
- 39 Bei M, Maas R. FGFs and BMP4 induce both Msx1-independent and Msx1-dependent signaling pathways in early tooth development. *Development* 1998; **125**:4325-4333.
- 40 Evseenko D, Zhu Y, Schenke-Layland K, et al. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proc Natl Acad Sci USA* 2010; **107**:13742-13747.
- 41 Mendjan S, Mascetti VL, Ortmann D, et al. NANOG and CDX2 pattern distinct subtypes of human mesoderm during exit from pluripotency. *Cell Stem Cell* 2014; **15**:310-325.
- 42 Yu P, Pan G, Yu J, Thomson JA. FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation. *Cell Stem Cell* 2011; **8**:326-334.
- 43 Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 2013; **8**:2281-2308.
- 44 Fu H, Ishii M, Gu Y, Maxson R. Conditional alleles of Msx1 and Msx2. *Genesis* 2007; **45**:477-481.
- 45 Nieminen P, Arte S, Pirinen S, Peltonen L, Thesleff I. Gene defect in hypodontia: exclusion of MSX1 and MSX2 as candidate genes. *Hum Genet* 1995; **96**:305-308.
- 46 Vieux-Rochas M, Bouhali K, Mantero S, et al. BMP-mediated functional cooperation between Dlx5;Dlx6 and Msx1;Msx2 during mammalian limb development. *PLoS One* 2013; **8**:e51700.
- 47 Qadir AS, Lee HL, Baek KH, et al. Msx2 is required for TNF-alpha-induced canonical Wnt signaling in 3T3-L1 preadipocytes. *Biochem Biophys Res Commun* 2011; **408**:399-404.
- 48 Bernardo AS, Faial T, Gardner L, et al. BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* 2011; **9**:144-155.
- 49 Brugger SM, Merrill AE, Torres-Vazquez J, et al. A phylogenetically conserved cis-regulatory module in the Msx2 promoter is sufficient for BMP-dependent transcription in murine and Drosophila embryos. *Development* 2004; **131**:5153-5165.
- 50 Sun J, Ishii M, Ting MC, Maxson R. Foxc1 controls the growth of the murine frontal bone rudiment by direct regulation of a Bmp response threshold of Msx2. *Development* 2013; **140**:1034-1044.
- 51 Hussein SM, Duff EK, Sirard C. Smad4 and beta-catenin co-activators functionally interact with lymphoid-enhancing factor to regulate graded expression of Msx2. *J Biol Chem* 2003; **278**:48805-48814.
- 52 Fong H, Hohenstein KA, Donovan PJ. Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. *Stem Cell* 2008; **26**:1931-1938.
- 53 Graham V, Khudyakov J, Ellis P, Pevny L. SOX2 functions to maintain neural progenitor identity. *Neuron* 2003; **39**:749-765.
- 54 Wang Z, Oron E, Nelson B, Razis S, Ivanova N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* 2012; **10**:440-454.
- 55 Lin IY, Chiu FL, Yeang CH, et al. Suppression of the SOX2 neural effector gene by PRDM1 promotes human germ cell fate in embryonic stem cells. *Stem Cell Rep* 2014; **2**:189-204.
- 56 Lengler J, Bittner T, Munster D, Gawad Ael D, Graw J. Agonistic and antagonistic action of AP2, Msx2, Pax6, Prox1 AND Six3 in the regulation of Sox2 expression. *Ophthalmic Res* 2005; **37**:301-309.
- 57 Tomioka M, Nishimoto M, Miyagi S, et al. Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res* 2002; **30**:3202-3213.
- 58 Twizere JC, Lefebvre L, Collete D, et al. The homeobox protein MSX2 interacts with tax oncoproteins and represses their transactivation activity. *J Biol Chem* 2005; **280**:29804-29811.
- 59 Yoon WJ, Cho YD, Cho KH, et al. The Boston-type craniosynostosis mutation MSX2 (P148H) results in enhanced susceptibility of MSX2 to ubiquitin-dependent degradation. *J Biol Chem* 2008; **283**:32751-32761.
- 60 Arnold SJ, Robertson EJ. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat Rev Mol Cell Biol* 2009; **10**:91-103.
- 61 Schier AF, Shen MM. Nodal signalling in vertebrate development. *Nature* 2000; **403**:385-389.
- 62 Sumi T, Tsuneyoshi N, Nakatsuji N, Suemori H. Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. *Development* 2008; **135**:2969-2979.
- 63 Lallemand Y, Nicola MA, Ramos C, Bach A, Cloment CS, Robert B. Analysis of Msx1; Msx2 double mutants reveals multiple roles for Msx genes in limb development. *Development* 2005; **132**:3003-3014.
- 64 Zhang X, Huang CT, Chen J, et al. Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* 2010; **7**:90-100.
- 65 Suwinska A, Czolowska R, Ozdzanski W, Tarkowski AK. Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. *Dev Biol* 2008; **322**:133-144.
- 66 Bessonard S, De Mot L, Gonze D, et al. Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* 2014; **141**:3637-3648.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)